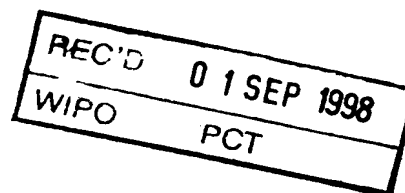




09/463480

PCT/AU98/00587



Patent Office  
Canberra

I, KIM MARSHALL, MANAGER EXAMINATION SUPPORT AND SALES, hereby certify that the annexed is a true copy of the Provisional specification in connection with Application No. PO 8233 for a patent by THE UNIVERSITY OF MELBOURNE filed on 25 July 1997.

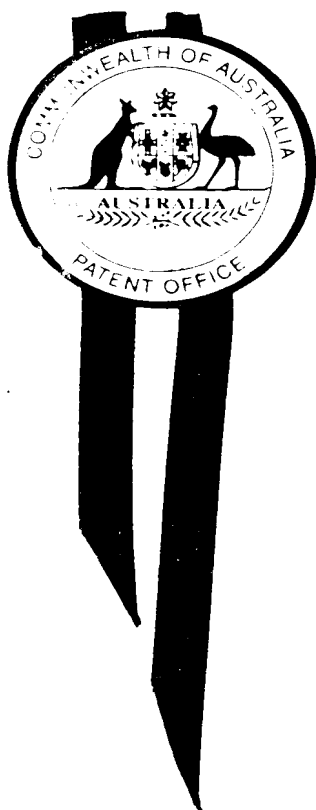
I further certify that the annexed specification is not, as yet, open to public inspection.

## PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1(a) OR (b)

WITNESS my hand this Third  
day of August 1998

KIM MARSHALL  
MANAGER EXAMINATION SUPPORT AND  
SALES



The University of Melbourne

**A U S T R A L I A**

**Patents Act 1990**

**PROVISIONAL SPECIFICATION**

for the invention entitled:

"Novel nucleic acid molecules and uses therefor"

The invention is described in the following statement:

- 1A -

## NOVEL NUCLEIC ACID MOLECULES AND USES THEREFOR

The present invention relates generally to a novel nucleic acid molecule. More particularly, the present invention relates to a male germ line cell specific genetic sequence in plants. Even more particularly, the present invention provides a male germ line specific gene or functional equivalent thereof and to the promoter of said gene or its functional derivatives and there use in generating a range of mutant plants including male sterile plants and transposon tagged plants.

Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element or integer or group of elements or integers but not the exclusion of any other element or integer or group of elements or integers.

Bibliographic details of the publications numerically referred to in this specification are collected at the end of the description. Sequence Identity Numbers (SEQ ID NOs.) for the nucleotide and amino acid sequences referred to in the specification are defined following the bibliography.

The increasing sophistication of recombinant DNA technology is greatly facilitating research and development in a range of industries and is particularly beneficial for the agricultural and horticultural industries. The ability to manipulate plants and plant products by recombinant means offers great potential to generate relatively quickly new varieties of plants, plants with beneficial genetic alterations and modified plant products, such as grains and fruits.

One important area of the plant industry is the production of hybrid plants. The production of hybrid plants from essentially homozygous parents permits the introduction of a range of beneficial traits including disease resistance, higher seed yield, frost resistance and altered nutritional characteristics.

Due to the importance of hybrid plants to the agricultural and horticultural industries in general, much research has been undertaken to finding improved, more efficacious ways of producing heterozygotic plants. The production of hybrid plants requires that a female parent does not self-

- 2 -

fertilize. A range of physical, chemical and genetic techniques have been used or proposed to prevent self-fertilization. Although some of these techniques have been partially successful, there is still a need to develop alternative, more broadly applicable methods of preventing self-fertilization.

5

Another important area of the agricultural and horticultural industries is the generation of mutants. Mutant plants may in themselves be useful in removing unwanted traits or may be useful as recipients for further genetic manipulation such as the introduction of new genetic material. Mutant plants have been obtained by a range of procedures including chemical and  
10 genetic manipulation as well as physical manipulation and classical breeding. One particularly useful mutant generating mechanism is "transposon tagging".

Transposons are distinct genetic elements capable of inserting into different sites of the genome within the same cell. Two broad categories of transposons are known comprising the DNA  
15 based transposon which transpose *via* DNA intermediates and retrotransposons or retroelements, which transpose *via* RNA intermediates. Transposons are useful tools for transposon tagging which relies upon a recognizable phenotype being caused by the insertion into a gene of a transposon. Transposon tagging has found particular application in the cloning of genes.

20 One system of transposon tagging uses the *Activator/Dissociation (Ac/Ds)* elements from maize (1). This system comprises a *trans*-activator, *Ac*<sup>Δ</sup>, which provides a transposase and a *cis*-responsive *Ds* element. The transposase promotes high frequency germinal excision of *Ds* which then reintegrates frequently into new genomic sites after excision.

25 However, despite the need for male sterile plants and the availability of mutagenic techniques such as transposon tagging, progress has been hampered by the inability to target germ line cells. In work leading up to the present invention, the inventors have identified cDNA clones exhibiting strict generative cell specific expression.

30 The development of male gametes is one of the most important events in the life cycle of flowering plants. The generative cell, the progenitor of male gametes, plays a central role in this

- 3 -

process. This role is to produce two male gametes, the sperm cells, which participate in fertilization. The generative cell resides within the cytoplasm of another cell, the vegetative cell and, until now, was thought to be transcriptionally inactive.

- 5 In work leading up to the present invention, the inventors have identified genes which are male gamete specific. The genes and their corresponding promoters of the present invention will enable specific genetic manipulation of the male germ line including generating male sterile plants and facilitating male gamete specific transposon tagging.
- 10 Accordingly, one aspect of the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence or a complementary sequence corresponding to a gene or derivative thereof or a regulatory region facilitating expression of said gene wherein said gene is specifically expressed in a male gamete of a plant.
- 15 The nucleic acid molecule of the present invention extends to a genomic or cDNA molecule corresponding to a gene or its derivative or a promoter of said gene or a functional derivative of said promoter, provided the promoter permits male gamete specific expression of the gene or its derivative.
- 20 The plant may be a monocotyledonous or dicotyledonous plant. Preferred plants include but are not limited to legumes, crop, cereal and native grasses, fruiting plants, flowering plants amongst many others.

In another embodiment, the present invention is directed to a nucleic molecule comprising a  
25 nucleotide sequence or complementary sequence encoding an amino acid sequence selected from SEQ ID NO:4, SEQ ID NO:6 and SEQ ID NO:8 or an amino acid sequence having at least 40% similarity to any one of SEQ ID NO:4, SEQ ID NO:6 or SEQ ID NO:8 wherein said nucleic acid molecule exhibits male gamete specific expression in plants.

- 30 Preferably, the percentage similarity is at least about 50%, more preferably at least about 60%, still more preferably at least about 70%, yet even more preferably at least about 80-90% or

- 4 -

greater to any one of SEQ ID NO:4, SEQ ID NO:6 or SEQ ID NO:8.

Another aspect of the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence or complementary nucleotide sequence selected from the group consisting  
5 of SEQ ID NO:3, SEQ ID NO:5 and SEQ ID NO:7 or a nucleotide sequence having at least 50% similarity to any one of SEQ ID NO:3, SEQ ID NO:5 or SEQ ID NO:7 or is a nucleotide sequence capable of hybridizing to any one of SEQ ID NO:3, SEQ ID NO:5 or SEQ ID NO:7 under low stringency conditions at 42°C.

10 Preferably, the percentage level of nucleotide similarity is at least about 60%, more preferably at least about 70%, still more preferably at least about 80%, yet still more preferably at least about 90% or greater to any one of SEQ ID NO:3, SEQ ID NO:5 or SEQ ID NO:7.

Reference herein to a low stringency at 42°C includes and encompasses from at least about 1%  
15 v/v to at least about 15% v/v formamide and from at least about 1M to at least about 2M salt for hybridisation, and at least about 1M to at least about 2M salt for washing conditions. Alternative stringency conditions may be applied where necessary, such as medium stringency, which includes and encompasses from at least about 16% v/v to at least about 30% v/v formamide and from at least about 0.5M to at least about 0.9M salt for hybridisation, and at least about 0.5M  
20 to at least about 0.9M salt for washing conditions, or high stringency, which includes and encompasses from at least about 31% v/v to at least about 50% v/v formamide and from at least about 0.01M to at least about 0.15M salt for hybridisation, and at least about 0.01M to at least about 0.15M salt for washing conditions.

25 Reference to a "derivative" herein includes single or multiple nucleotide or amino acid substitutions, deletions and/or additions as well as parts, fragments, portions, homologues and analogues of the nucleotide or amino acid sequence.

The nucleic acid molecules of the present invention are specifically expressed in male gametes  
30 of plants, ie. in the generative cells. The male gamete specific expression is determined in part by the male gamete specific promoter.

- 5 -

Accordingly, another aspect of the present invention provides a nucleic acid molecule comprising a promoter or functional derivative thereof which directs plant male gamete specific expression in a nucleotide sequence operably linked thereto.

- 5 More particularly, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence or complementary nucleotide sequence which is capable of hybridizing under low stringency conditions at 42°C to a genomic region encompassing at least about 2kbp upstream of the nucleotide sequence corresponding to any one of SEQ ID NO:3 or SEQ ID NO:5 or SEQ ID NO:7 and wherein said nucleic acid molecule is capable of directing plant male  
10 gamete specific expression of a nucleotide sequence operably linked thereto.

The identification of the male gamete specific promoters and genes permits the generation of a range of male sterile plants as well as male gamete specific transposon tagging.

- 15 In one embodiment, the present invention contemplates a method of inducing or otherwise facilitating male sterility in a plant, said method comprising operably linking a cytotoxic nucleic acid molecule to a promoter which directs male gamete specific expression in said plant such that upon expression of said promoter, the cytotoxic nucleic acid molecule is expressed to produce a product which inactivates, kills or otherwise renders substantially non-functional male gametes  
20 in said plant.

The cytotoxic nucleic acid molecule may encode or comprise a cytotoxic protein, an antisense molecule to a particular gene, a ribozyme or a plantabody amongst many other molecules.

- 25 Preferably, the promoter corresponds to a nucleotide sequence which hybridizes under low stringency conditions to a genomic region comprising at least about 2kbp upstream of a gene corresponding to any one of SEQ ID NO:3, SEQ ID NO:5 or SEQ ID NO:7.

Alternatively, the cytotoxic nucleic acid molecule is fused to the gene naturally operably linked  
30 to said promoter such that upon expression of said gene, the cytotoxic nucleic acid molecule inactivates, kills or otherwise renders substantially non-function a male gamete in said plant.

In another embodiment, the male gamete specific promoter and/or gene is used to facilitate male gamete specific transposon tagging. This facilitates the product of pollen grains in a plant carrying a transposon tag. Offspring can then be screened for a range of phenotypes of interest and then, in turn, the transposon tagged plants used to clone particular genes.

5

Accordingly, another aspect of the present invention provides a genetic construct comprising a male gamete specific promoter, as hereinbefore described, operably linked to a transposase gene, said transposase gene capable of inducing transposition of a transposable element, such that upon expression of said promoter, the transposase gene is expressed facilitating transposition of said  
10 transposable element.

A particularly useful transposon system is the Ds<sup>ALS</sup> system (1, 5) where the activator (Ac) transposase would be under the control of the promoter of the present invention to facilitate transposition of the dissociation (Ds) element.

15

In accordance with the present invention a plant is selected such as a crop plant, legume, grass plant or flowering plant amongst other monocots and dicots and a callus culture prepared. A genetic construct comprising the male gamete specific promoter and optionally male gene specific gene naturally associated with said promoter operably linked to a cytotoxic nucleic acid  
20 molecule or a transposase gene is introduced into callus cells. A plant is then regenerated. The male gamete specific construct may be under additional control mechanisms such as environmental, developmental, physiological or nutritional control mechanisms such that upon provision of these mechanisms, the male gamete specific promoter is activated. In any event, upon expression of the male gamete specific promoter, transposon tagging will occur or the  
25 cytotoxic nucleic acid will be expressed. This will result in tagged pollen or male sterility.

Male sterile plants containing a range of transposon insertions and genetic constructs useful of the practice of the present invention are all encompassed by the present invention as are all offspring or progeny, new plant varieties and mutant plants.

30

The present invention extends to the promoter as herein described as well as functional mutants



- 7 -

thereof. A functional mutant includes promoter fusions to other promoters, as well as single or multiple nucleotides, deletions, additions and/or substitutions including parts, fragments, portions, homologues and analogues thereof.

- 5 Although not intending to limit the present invention to any one type of male gamete specific gene or promoter, genes and their promoters encoding histones are particularly useful.

Another benefit of the present invention provides the potential to develop seedless fruit or fruit with reduced seed content. This is particularly applicable where pollination stimulates fruit  
10 development and where the lack of fertilization results in seedless fruit.

The present invention extends to any transposable element such as but not limited to *Ac*, *Ds*, *En/Spm*, *dspm*, *Tam3*, *dTam3*, *Mu1*, *Tat1*, *Tag1*, *dTph1*, *Tnt1*, *Tto1*, *Tto2*, *Ac-like*, *dTnp* and *Tos17*. These elements are conveniently reviewed in the reference (16).

15

The present invention is further described by the following non-limiting Figures and/or Examples.

**Figure 1** is a representation of the nucleotide [SEQ ID NO:3] and predicted amino acid [SEQ ID NO:4] sequence of *LGC1*.

20

**Figure 2** is a photographic representation showing expression of *LGC1* mRNA in different tissues of lily. (A) Northern blot of the indicated tissues probed with <sup>32</sup>P-labelled *LGC1* probe. GCs, generative cells. (B) RT-PCR of different tissues. Pollen mRNA includes contribution of both generative cell and vegetative cell. Numbers 16, 31, 64 represent 1/16, 1/32, and 1/64 of  
25 the mRNA input respectively and so forth. Molecular sizes are indicated on the left.

**Figure 3** is a photographic representation showing in situ hybridization of *LGC1* mRNA to whole mount lily pollen. Dark staining in the generative cell (arrowhead) represents hybridization signals detected by an alkaline phosphatase conjugated anti-DIG antibody. The  
30 outer wall of pollen, exine appears as a sculptured pattern. (A) Pollen probed with a DIG-UTP labelled *LGC1* antisense riboprobe. (B) Control pollen probed with a sense riboprobe.

**Figure 4** is a photographic representation showing *in situ* hybridization of *LGC1* mRNA to whole mount lily pollen at different developmental stages. For a better resolution, protoplasts of developing pollen were released from sculptured exine, the outer wall of pollen (9). Developing pollen (A-E) and pollen tube (K) probed with a DIG-UTP labelled riboprobe and then counter-stained with 4', 6'-diamidino-2-phenyl indole (DAPI) to visualize the vegetative and generative nuclei within pollen (F-J) and sperm nuclei in pollen tube (L). Arrowheads indicate the generative cell at early developmental stages. GN, generative nucleus; VN, vegetative nucleus; SC, sperm cell; SN, sperm nucleus.

**Figure 5** is a representation showing nucleotide [SEQ ID NO:5] and deduced amino acid [SEQ ID NO:6] sequences of the *gcH2A* cDNA. The predicted amino acid sequence (numbered at right) is given below the corresponding nucleic acid sequence (numbered at left).

**Figure 6** is a representation showing nucleotide [SEQ ID NO:7] and deduced amino acid [SEQ ID NO:8] sequences of the Full Length *gcH3* cDNA. Numbers at left indicate base positions of the nucleotide sequence, numbers at right residue positions of the derived amino acid sequence.

**Figure 7** is a photographic representation showing expression pattern of *gcH2A* and *gcH3*.

**Figure 8** is a photographic representation showing *in situ* hybridization of *gcH2A* and *gcH3* in pollen. Pollen exine was removed for a better visualising of signal.

(A) Pollen probed with showing strong hybridization signal in the generative cell.

(B) Control pollen probed with DIG-labelled sense *gcH2A*.

(C) Pollen probed showing strong hybridization signal in the generative cell.

(D) Control pollen probed with DIG-labelled sense *gcH3*.

**Figure 9** is a photographic representation showing expression of *gcH2A* and *gcH3* during pollen development. *In situ* hybridization of microspores immediately after formation of generative cell (A, D, G), nearly mature pollen (B, E, H) and mature pollen (C, F, I). Arrow heads indicate nearly formed generative cell, VN, vegetative nucleus, GN, generative cell nucleus. Pollen exine was removed for a better visualising of signal.

- 9 -

(A), (B), (C) samples probed with DIG-labelled antisense gcH2A showing strong hybridization signal only in mature pollen.

(G), (H), (I) samples probed with DIG-labelled antisense gcH3 showing hybridization signal only in mature pollen.

5 (D), (E), (F) DAPI staining of corresponding developmental stages.

- 10 -

### EXAMPLE 1

#### ISOLATION OF *LGCI*

Generative cells from lily (*Lilium longiflorum*) were isolated and mRNA isolated therefrom.  
5 Generative cells were isolated from fresh pollen of lily as previously described (6) and stored at  
-70°C until use. mRNA was extracted directly from approximately  $1 \times 10^5$  of stored generative  
cells using a mRNA purification kit (Pharmacia-LKB). Purified generative cell mRNA was  
reverse transcribed and the resultant cDNA was amplified by PCR, size fractionated and cloned  
into  $\lambda$ gt11 expression vector.

10

A differential hybridization approach was used to obtain a cDNA clone corresponding to a gene  
specifically expressed in generative cells. The clone was designated *LGCI*. In the differential  
hybridization approach, a number of cDNA clones were randomly picked from a generative cell  
cDNA library and cDNA inserts obtained by PCR with  $\lambda$ gt11 forward and reverse primers. PCR  
15 conditions were 30 cycles of 1 min at 94°C, 2 min at 60°C and 3 min at 72°C with a final  
extension at 72°C for 10 min. The amplified cDNA inserts were purified, labelled with  $^{32}\text{P}$  by  
random priming (Bresatec Ltd, South Australia) and used for probing of RNA slot blots  
containing approximately 300 ng of mRNAs from various tissues including leaf, stem, petal,  
stigma/style, ovary, pollen and generative cells. Hybridization and washing was performed as  
20 previously described (18). cDNA clones showing preferential or specific hybridization to  
generative cell mRNA were selected for further analysis.

The cDNA insert of one clone, *LGCI*, was subcloned into pBluescript(SK)+(Stratagene) and  
sequenced with ABI PRISM (trademark) dye terminator cycle sequencing kit (Perkin-Elmer).  
25 The *LGCI* cDNA insert was shown to be 618 bp in length encoding a predicted gene product  
of 128 amino acids with a calculated molecular weight of 13.8 kDa (Figure 1). *LGCI*  
corresponds to a 0.6 kbp transcript which is present at a high level in generative cells as revealed  
by Northern blot analysis (Figure 2A).

30 No signal was detectable in the two vegetative tissues tested, leaf and stem, while a faint signal  
was visible in pollen containing generative cells. The tissue specificity of *LGCI* was further

- 11 -

examined by RT-PCR using gene specific PCR primers that amplify a 0.3 kbp portion of the coding region. For RT-PCR, mRNAs from generative cells and various tissues were reverse transcribed and amplified by PCR with a pair of sequence specific primers (L13A: 5'-GTACTCTTAAGCATACAACATGAG -3' [SEQ ID NO:1]; L13B: 5'-  
5 CAGGCATACTTGAATGCTACAAGA-3' [SEQ ID NO:2]) using the Access RT-PCR System (Promega). For each tissue, mRNA was subjected to a serial two-fold dilutions. Based on the signal intensity of the amplified products, the relative amount of *LGCI* mRNA in each tissue was estimated.

10 RT-PCR amplifications were performed using controlled amount of RNA input from various tissues of lily plant. A PCR product of expected size (0.3 kbp) was obtained in generative cells and pollen but not in all the other tissues tested including vegetative parts such as leaf, stem as well as reproductive parts such as petal, female stigma/style and ovary (Figure 2B). Based on the signal intensity, the inventors estimated that approximately 20 fold more PCR product was  
15 obtained when generative cell mRNA was used as compared to pollen mRNA. Since the generative cell constitutes a small portion of pollen, the inventors considered that the amplified *LGCI* product obtained using pollen mRNA input may represent the contribution of generative cell only. Generative cell specificity of *LGCI* was further confirmed by *in situ* hybridization as hereinafter described.

20

Non-radioactive whole mount *in situ* hybridization was performed in both developing and mature pollen based on the protocols previously described (3, 4, 5). Fresh pollen at various developmental stages was fixed (1% v/v glutaraldehyde in 50 mM PIPES buffer, pH 7.4) for 2 hours at room temperature. The fixed pollen was then washed in buffer and stored in 70% v/v  
25 ethanol at 4°C until use. Both sense and antisense riboprobes labelled with DIG-UTP were generated from linearized DNA templates. The hybridization signal was detected with an alkaline phosphatase conjugated anti-DIG antibody using a DIG nucleic acid detection kit (Boehringer Mannheim). To obtain a better resolution, protoplasts of developing pollen were released from exine (the outer wall of pollen) by treatment with enzyme solution (1% w/v  
30 Macerozyme, 0.5% w/v Cellulase and 0.5% w/v BSA) as previously described (6). Vegetative and generative nuclei within pollen were visualized by counter-staining with 4', 6'-diamidino-2-

phenyl indole (DAPI).

The results clearly showed that *LGC1* mRNA is confined to the generative cell in mature pollen (Figure 3). *LGC1* mRNA in pollen as detected by Northern blot and RT-PCR own their origin  
5 to the generative cell.

To determine whether *LGC1* mRNA present in the generative cell is the product of generative cell specific gene activity or the result of asymmetric RNA localization and partitioning prior to generative cell formation in developing pollen, the inventors monitored *LGC1* mRNA  
10 accumulation during this process. The inventors examined six different developmental stages of generative cells. At the early stage, the newly formed generative cell is attached at one pole of pollen with the vegetative nucleus located in its vicinity (Figures 4A, F). As the development progresses, the generative cell starts to detach itself from the intine (inner cell wall of pollen) while the vegetative nucleus moves towards the centre of pollen (Figures 4B, G). No detectable  
15 signal was observed in these two early developmental stages (Figures 4A, B). With rapid size expansion of pollen, the generative cell separates completely from the intine and suspends freely within the vegetative cell cytoplasm. At this stage, its shape becomes elongated with a large nucleus in the centre and most of cytoplasm at both ends of the cell (Figures 4C, H). A weak signal was detected at both ends of the generative cell, indicating the initiation of *LGC1* mRNA  
20 transcription (Figures 4C). As the development continues, the generative cell becomes spindle-shaped (Figures 4D, I) and accumulation of *LGC1* mRNA in the generative cell becomes more evident (Figures 4D). At the time of pollen maturity, a very high level of *LGC1* mRNA were observed in the generative cell (Figure 3A, Figures 4E, J). Next, pollen germination occurs on female stigma and pollen tubes grow inside the female stylar tissue. The generative cell then  
25 moves into pollen tube and undergoes a mitotic division producing two male gametes, the sperm cells (Figures 4K, L). *LGC1* mRNA was clearly detectable in the two sperm cells inside the pollen tubes (Fig. 4K) as described more fully below.

In lily, generative cell division occurs in the pollen tube during its growth in the female stylar  
30 tissue. *In situ* hybridization of mRNA in sperm cells, therefore, can only be performed in pollen tube. Pollen tubes were grown *in vivo* by hand pollinating pistils with freshly collected pollen.

- 13 -

After 48 hours, a 1 cm long segment was taken from the base of the style and cut into two symmetrical halves. Pollen tubes growing in the hollow stylar canal were teased out, fixed and then used for *in situ* hybridization as described above.

- 5 No signal was detected in the vegetative cell at any stage of pollen development. These results show that the generative cell specific accumulation of *LGC1* mRNA is due to differential gene activation of generative cell.

Male germ line specific gene expression represents a new aspect of fundamental importance in  
10 flowering plants. *LGC1* is the first male germ line specific gene to be identified in flowering plants and thus, the present study of generative cell specific gene expression has important implications in understanding the molecular bases of male gamete development. Several aspects of research can immediately benefit from the availability of this gene and its promoter. For example, selective ablation of the male gametes can be achieved using generative cell specific  
15 promoter- cytotoxin fusions. The availability of *LGC1* gene promoter will make it possible to introduce marker genes for monitoring the process of sperm-egg recognition and fusion at molecular level. Furthermore, the male gamete specific promoter may be used to generate a range of transposons to specify tagged pollen genes.

20

## EXAMPLE 2

### MALE GAMETE CELL SPECIFIC EXPRESSION OF H2A AND H3 HISTONE GENES

The following Examples shows the identification of two cDNA clones, *gcH2A* and *gcH3*, which  
25 encode male gamete-specific variants of histones H2A and H3, respectively. The inventors show that both *gcH2A* and *gcH3* mRNAs accumulate exclusively within the male germ line cell, the generative cell. An examination of the spatial distribution of *gcH2A* and *gcH3* transcripts during pollen development show that initiation of expression of these genes occurs in generative cell at the later stages of pollen maturation. The results indicate that these histone variants are the  
30 products of generative cell transcriptional activity. This example provides the first insight of male germ line cell specific histone gene expression in flowering plants.

## 1. INTRODUCTION

Histones are the major protein constituents of the chromatin of eukaryotic cell nuclei. Histone proteins include five major classes: four core histones, H2A, H2B, H3, H4 and one linker histone H1. The core histones are small, basic proteins (11-15 kDa) that contain a high proportion of positively charged amino acids, mainly lysine and arginine. Histones are highly conserved throughout evolution and are encoded by multigene families. Genes encoding major classes of histones are usually expressed in a cell cycle-dependent fashion at the beginning of the S (DNA synthesis) phase and are co-ordinately regulated at the transcriptional and post-transcriptional level through the cell cycle (7).

## 2. METHODS

### 15 (a) Construction and screening of cDNA library

Generative cells were isolated from mature pollen of lily (*Lilium longiflorum*) as previously described (8) and stored at -70°C until use. Poly(A)+ RNA was isolated from approximately  $1 \times 10^5$  of stored generative cells using oligo (dT)-cellulose affinity column (Pharmacia) according to the manufacture's instruction. First-strand cDNA was synthesized with an oligo (dT) primer. A Capswitch primer was also used to ensure the synthesis of full length clones. The resultant cDNA was amplified by PCR using the following conditions: 35 cycles of 94°C for 1 min, 42°C for 2 min and 72°C for 2 min. The PCR products were size-fractionated through a Sephadex-50 column and cDNAs of appropriate size were cloned into  $\lambda$ gt11 expression vector.

25

For screening, a number of cDNA clones was randomly picked and cDNA inserts were obtained by PCR with  $\lambda$ gt11 forward and reverse primers. Differential screening was conducted by probing RNA slot blots of various tissues with the amplified cDNA inserts. cDNA clones showing strong hybridization to generative cell RNA, weak hybridization to pollen RNA and no hybridization to other tissues were considered to be putative generative cell-specific clones.

30



**(b) Sequencing analysis**

The putative generative cell cDNA clones were subcloned into pBluescript II SK+ (Stratagene). Sequencing was performed on both strands by the dideoxy chain-termination method (9) using  
5 ABI PRISM (trademark) dye terminator cycle sequencing kit (Perkin-Elmer) with an automated DNA sequencer. Sequence-specific primers were used to generate overlapping sequence information. DNA and protein sequence analysis was performed using BLAST search tools.

**(c) RNA gel blot analyses**

10

Total RNA was prepared from various tissues (10). Generative cell RNA was isolated using SNAP RNA extraction kit (Invitro Gene) according to the manufacture's procedure. For gel blot analysis, 20  $\mu$ g of total RNA was separated by denatured agarose gel electrophoresis, blotted onto Hybond N+ nylon membrane (Amersham) and probed with <sup>32</sup>P-labelled *gcH2A* and *gcH3*  
15 cDNA inserts. Hybridization of probes with RNA blots was performed in 50% v/v deionised formamide, 2 x SSPE (1 x SSPE is 0.15 M NaCl, 0.01 M NaH<sub>2</sub>PO<sub>4</sub>, and 1 mM EDTA, pH 7.4), 1% w/v PEG, 0.5% w/v BLOTTO, 7% w/v SDS and 0.5mg/ml denatured salmon sperm DNA at 42°C overnight. The blots were washed with 2 x SSC (1 X SSC is 0.15 M NaCl and 15 mM sodium citrate, pH 7.0), 0.1% w/v SDS at room temperature for 15 min and with 0.2 x SSC,  
20 0.1% w/v SDS at 65°C for 15 min, followed by a brief wash in 0.2 x SSC. The blots were re-probed with lily ribosome RNA to verify the relative amount of RNAs loaded.

**(d) *In situ* hybridization**

25 Non-radioactive whole mount *in situ* hybridization was performed based on the protocols described (11, 12, 13). Developmental stages of pollen were determined using 4', 6'-diamidino-2-phenyl indole (DAPI) staining. Mature and developing pollen was treated with an enzyme solution (1% w/v macerozyme, 0.5% w/v cellulase and 0.5% w/v BSA) for 1 hour to remove the exine (the outer wall of pollen). Pollen protoplasts were then washed in 50 mM PIPES  
30 buffer and fixed in 1% v/v glutaraldehyde in 50 mM PIPES buffer, pH 7.4, for 2 hours at room temperature. The fixed pollen was then washed in 50 mM PIPES buffer and stored in 70% v/v

ethanol at 4°C.

Prior to hybridization, pollen samples were first dehydrated through an ethanol series up to 100% v/v ethanol. Samples were then treated with xylene (2 x 10 min) followed by rehydration through an ethanol series. Proteinase K (1 µg/ml) treatment was carried out in 100 mM Tris-HCl, pH 8 and 50 mM EDTA for 40 min at 37°C. Digoxigenin-labelled riboprobes were synthesized by *in vitro* transcription (Promega). Hybridization was performed in 50% v/v formamide, 6 x SSC, 3% w/v SDS, 100 µg/ml tRNA at 55°C overnight. Samples were then washed in 1 x SSC, 0.1% w/v SDS at room temperature followed by 2 x 10 min washes in 0.2 SSC, 0.1% w/v SDS at 55°C. RNase A (10 µg/ml) treatment was performed in 2 x SSC for 1 hour at 37°C. Hybridization signal was detected using a DIG detection kit (Boehringer Mannheim) according to the manufacture's specification. Vegetative and generative cell nuclei were visualized by counter-staining with DAPI.

## 15 RESULTS

### Isolation and Characterisation of histone *gcH2A* and *gcH3* cDNA clones

Lily (*Lilum longiflorum*) was used as an experimental system in accordance with the present Example. Within the pollen grain, the male germ line cell (generative cell) is enclosed in the much larger vegetative cell. To maximize the chance of obtaining genes specifically expressed in the generative cell, the inventors prepared a cDNA library using polyA(+) RNA from isolated generative cells. The cDNA library was screened by differential hybridization using probes from generative cells, pollen, leaf, stem, pistil and ovary. cDNA clones that gave strong positive hybridization signal with generative cell mRNA, weak signal with pollen mRNA and no signal with mRNA from other tissues were considered as putative generative cell specific clones. These cDNA clones were subjected to further analysis. Two of these clones were found to encode proteins which were identified as variants of histone H2A and H3, respectively. The two clones were designated "*gcH2A*" and "*gcH3*".

30

*gcH2A* cDNA is 581 bp long and contains an open reading frame of 333 bp starting from the first

ATG at position 49 to a stop codon TAA at position 379 (Figure 1). The derived amino acid sequence of *gcH2A* is composed of 111 amino acids and encodes a protein with a calculated molecular mass of 12.1 kDa. *gcH2A* polypeptide contains 10.8% arginine and 5.4 % lysine. The deduced amino acid sequence of *gcH2A* shows high levels of sequence similarity as well as  
5 variability when compared to somatic H2A histones from other organisms. The N-terminal region of the protein appeared to be more conserved than the C-terminal region. In addition, *gcH2A* polypeptide is 30-35 amino acids shorter at the C-terminus than somatic H2A histone. It has been reported that the C-terminal variable regions of wheat somatic histones can be of two structural different types (14). Type 1 H2A proteins have one or two copies of a SPKK motif  
10 which is known to interact with the minor groove of the DNA, whereas type 2 H2A proteins have a shorter C-terminal variable region and no SPKK motif. Using these criteria, the lily generative cell specific H2A (*gcH2A*) histone can be classified as type 2 since the C-terminal region of *gcH2A* does not contain a SPKK motif.

15 The complete sequence of the *gcH3* cDNA clone is shown in Figure 6. The *gcH3* cDNA is of 485 nucleotides and contains a putative open reading frame of 336 bp encoding a protein of 112 amino acids. The predicted *gcH3* polypeptide, containing 8% arginine and 12.5% lysine, has a calculated molecular mass of 12.5 kDa. When compared to somatic histone H3, the deduced amino acid sequence of *gcH3* exhibits two highly conserved regions located near both terminus  
20 of the polypeptide and a variable region of 14 amino acids (position 50 to 64) in the centre region.

Both *gcH2A* and *gcH3* histone clones were transcribed as polyadenylated mRNAs. Sequencing analysis revealed A/T rich regions resembling the polyadenylation consensus signal and  
25 polyadenylated tract bases at their 3' ends (Figures 5 and 6).

To determine the expression patterns of *gcH2A* and *gcH3*, RNA blot analysis was performed with RNA samples from various organs including generative cells, pollen grain, young expanding leaf, stem, pistil and ovary. Considering the highly conserved nature of the histone coding  
30 region, hybridization and washing were conducted at high stringency to avoid cross hybridizations with other somatic histone mRNAs. mRNAs corresponding to both *gcH2A* and

*gcH3* were detected in generative cells (Fig. 7). A weak hybridization signal was also detected in pollen whereas neither vegetative nor other floral tissues tested showed detectable levels of *gcH2A* and *gcH3* mRNAs. Since pollen grains contain both vegetative and generative cells, it was apparent that the fainter signal detected in pollen RNA was due to the contribution of  
5 generative cell only. The inventors tested young leaf and stem tissues from seedlings which have a large number of dividing cells by RNA gel blot as well as RT-PCR analyses. No expression, neither of *gcH2A* nor of *gcH3* was detected. Since the tissues tested represent a broad spectrum of plant organs, it was concluded that both *gcH2A* and *gcH3* are expressed in generative cells only. From the intensity of the hybridization signal, it can be assumed that *gcH2A* is a highly  
10 abundant gene, whereas *gcH3* represents a lowly expressed transcript.

The inventors examined the spatial distribution of *gcH2A* and *gcH3* mRNAs within pollen by *in situ* hybridization. Digoxigenin (DIG) labelled *gcH2A* and *gcH3* were used to probe whole-mount pollen grains. Accumulation of both *gcH2A* and *gcH3* mRNAs were clearly confined to  
15 the generative cell of pollen whereas no hybridization signal was detected in the vegetative cells of pollen (Figures 8a, c). No signal was observed in pollen grain probed with control sense probes (Figures 8b, d). The accumulation of *gcH2A* in the generative cell appeared much higher than that of *gcH3*. The results obtained by *in situ* hybridization correspond to those of RNA gel blot analysis and clearly demonstrate the generative cell specificity of both *gcH2A* and *gcH3*.

20

To determine the temporal expression of *gcH2A* and *gcH3*, the inventors examined five developmental stages of male gametogenesis. It is well established that three DNA replications occur during male gametogenesis of flowering plants. The first replication occurs prior to  
25 meiosis in the microsporocyte or pollen mother cell which produces a tetrad of four haploid microspores. The second replication occurs in the microspore before the first mitotic division (pollen mitosis I) which produces a large vegetative cell and a small generative cell. The third replication takes place in the generative cell before the second mitosis (pollen mitosis II) which results in the formation of two male gametes (sperm cells). To determine whether *gcH2A* and  
30 *gcH3* are associated with any of these three DNA replications during male gametogenesis, the inventors performed *in situ* hybridization in microsporocyte, microspore and three stages of

generative cell development. No hybridization signal was observed in pre-meiotic microsporocytes and pre-mitotic microspores. Further, no *gcH2A* and *gcH3* mRNAs were detected in the newly formed generative cell soon after pollen mitosis I (Figures 9a, d, g). As development progresses into pollen maturation, the generative cell completely separates from the intine wall of pollen and suspends freely within the vegetative cell cytoplasm. At this stage, the generative cell becomes elongated and spindle-shaped with a large nucleus in the centre and most of its cytoplasm at both ends (Figures 9b, e, h). A weak signal was observed at both ends of the generative cell when probing with *gcH2A*, indicating the initiation of *gcH2A* mRNA transcription (Figure 9b). At the time of pollen maturity, the accumulation of *gcH2A* mRNA in the generative cell reached a very high level as indicated by the strong hybridization signal (Figure 7c). In comparison to this, the signal obtained with *gcH3* probe appeared much weaker (Figure 7i), and mRNA corresponding to the *gcH3* clone could only be detected at the mature stage of pollen development.

Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations of any two or more of said steps or features.

## BIBLIOGRAPHY

1. Honma, M.A. *et al.* *Proc. Natl. Acad. Sci. USA* 90: 6242-6246, 1993.
2. Xu, H. *et al.* *Gene* 164: 255-259, 1995.
3. Bouget, F. *et al.* *J. Phycol.* 31: 1027-1030, 1995.
4. Bouget, F. *et al.* *Plant Cell* 8: 189-201, 1996.
5. Torres, M *et al.* *Plant J.* 8: 317-321, 1995.
6. Blomstedt, C.K. *et al.* *Plant Mol. Biol.* 31: 1083-1086, 1996.
7. Osley, M.A. *Annu. Rev. Biochem.* 60: 827- , 1991.
8. Tanaka, I. *Protoplasma* 142: 68-73, (1988)
9. Sanger, F. *et al.* *Proc. Natl. Acad. Sci. USA* 74: 5463-5467, 1977.
10. Chomczynski, P. *et al.* *Anal. Biochem.* 162: 156-159, 1987.
11. Bouget, F. *et al.* *J. Phycol.* 31: 1027-1030, 1995.
12. Terres, M.A. *et al.* *Plant J.* 8: 317-321, 1995.
13. Bouget, F. *et al.* *Plant Cell* 8: 189-201, 1996.
14. Huh, H.G. *et al.* *Plant Mol. Biol.* 33: 791-802, 1997.
15. Rommens, C.M.T. *et al.* *Mol. Gene. Genet.* 231: 433-441, 1992.
16. Roberts, M. In *Plant Gene Isolation: Principles and Practice*, Ed. by C.D. Foster and D. Twell, pp 301-328, 1996, John Wiley & Sons Ltd.
17. McCormick, S. *Plant Cell* 5: 1265-1275, 1993.

- 21 -

## SEQUENCE LISTING

### (1) GENERAL INFORMATION:

- (i) APPLICANT: THE UNIVERSITY OF MELBOURNE
- (ii) TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES THEREFOR

### (iii) NUMBER OF SEQUENCES: 8

### (iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: DAVIES COLLISON CAVE
- (B) STREET: 1 LITTLE COLLINS STREET
- (C) CITY: MELBOURNE
- (D) STATE: VICTORIA
- (E) COUNTRY: AUSTRALIA
- (F) ZIP: 3000

### (v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25

### (vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER: NEW PROVISIONAL
- (B) FILING DATE:
- (C) CLASSIFICATION:

### (viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: HUGHES, DR E JOHN L
- (C) REFERENCE/DOCKET NUMBER: EJH/AF

### (ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: +61 3 9254 2777
- (B) TELEFAX: +61 3 9254 2770
- (C) TELEX: AA 31787

- 22 -

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 14 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GTACTCTTAA GCATACAACA TGAG

14

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 14 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CAGGCATACT TGAATGCTAC AAGA

14

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 625 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 82..468

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GCCATCCCAT CAACAGAAGG TTAAAGTGG AATCCATTTC ATTAGAAAAG ATCGGACAAA	60
GGGTACTCTT AAGCATACAA C ATG AGG GCG GTG GCG GTT TTC TTT GCT TGC	111
Met Arg Ala Val Ala Val Phe Phe Ala Cys	
1 5 10	
GTT CTC TTC TGT ATG GTT CAC AAA GCC GCA CTT GCG GAT GAT AAA ACG	159
Val Leu Phe Cys Met Val His Lys Ala Ala Leu Ala Asp Asp Lys Thr	
15 20 25	
TGC AAC CCT ACA GAT TTT ATG GTT ACC CAA ACC ATA ACT GGA TTG ACA	207



- 23 -

Cys	Asn	Pro	Thr	Asp	Phe	Met	Val	Thr	Gln	Thr	Ile	Thr	Gly	Leu	Thr		
			30					35					40				
ATC	GGC	GGT	AAA	CAA	GAG	TTC	GAG	GTC	AAT	TTA	ATA	AAC	AAT	TTG	TAT		255
Ile	Gly	Gly	Lys	Gln	Glu	Phe	Glu	Val	Asn	Leu	Ile	Asn	Asn	Leu	Tyr		
		45					50					55					
TGT	GCA	CAA	TCT	AAT	GTC	AAA	GTT	TCA	TGT	GAC	GGG	CTT	CAT	ACC	ACC		303
Cys	Ala	Gln	Ser	Asn	Val	Lys	Val	Ser	Cys	Asp	Gly	Leu	His	Thr	Thr		
	60					65					70						
GAA	CCA	ATA	GAT	CCT	CAC	ATT	ATC	AGA	CCA	CTT	AGT	GAC	GGA	ACG	AAC		351
Glu	Pro	Ile	Asp	Pro	His	Ile	Ile	Arg	Pro	Leu	Ser	Asp	Gly	Thr	Asn		
	75				80					85					90		
AAC	TGC	CTT	GTC	AAC	AAT	GGA	GCG	CCT	ATT	TCT	CAT	GCT	ACT	CTT	GTA		399
Asn	Cys	Leu	Val	Asn	Asn	Gly	Ala	Pro	Ile	Ser	His	Ala	Thr	Leu	Val		
				95					100					105			
GCA	TTC	AAG	TAT	GCC	TGG	GAT	GTT	CCT	CCA	TCT	TTC	AGC	ATC	ATC	AGC		447
Ala	Phe	Lys	Tyr	Ala	Trp	Asp	Val	Pro	Pro	Ser	Phe	Ser	Ile	Ile	Ser		
			110					115					120				
TCT	GAT	ATA	AAT	TGC	TCC	TAA	GGAGAAA	ATTCTAGTTG	GCAGAGAATA								495
Ser	Asp	Ile	Asn	Cys	Ser	OCH											
		125															
ATCATATAGT	CTTTTTTACT	GAGCTATTTA	ATTTTTTTCAA	TTTTTCACCAA	TAAGATTATT												555
TTAATGGAAT	GTTAATGTAT	TAGAATTGAA	AAATAAAAAA	AAAAAAAAAA	AAAAAAAAAA												615
AAAAAAAAAA																	625

## (2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 128 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met	Arg	Ala	Val	Ala	Val	Phe	Phe	Ala	Cys	Val	Leu	Phe	Cys	Met	Val		
1				5					10					15			
His	Lys	Ala	Ala	Leu	Ala	Asp	Asp	Lys	Thr	Cys	Asn	Pro	Thr	Asp	Phe		
		20						25					30				
Met	Val	Thr	Gln	Thr	Ile	Thr	Gly	Leu	Thr	Ile	Gly	Gly	Lys	Gln	Glu		
		35					40					45					
Phe	Glu	Val	Asn	Leu	Ile	Asn	Asn	Leu	Tyr	Cys	Ala	Gln	Ser	Asn	Val		
	50					55					60						
Lys	Val	Ser	Cys	Asp	Gly	Leu	His	Thr	Thr	Glu	Pro	Ile	Asp	Pro	His		
	65				70					75				80			
Ile	Ile	Arg	Pro	Leu	Ser	Asp	Gly	Thr	Asn	Asn	Cys	Leu	Val	Asn	Asn		
			85						90					95			
Gly	Ala	Pro	Ile	Ser	His	Ala	Thr	Leu	Val	Ala	Phe	Lys	Tyr	Ala	Trp		
		100						105					110				
Asp	Val	Pro	Pro	Ser	Phe	Ser	Ile	Ile	Ser	Ser	Asp	Ile	Asn	Cys	Ser	OCH	

- 24 -

115

120

125

## (2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 587 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 49..378

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```

GAAAGTTGAA ACATCTCCAT CAAACTCTAG AGTCAGATTT CCCACAAG ATG ATT TCA      57
                                   Met Ile Ser
                                   1
TCG GCA AAT AAC AAA GGC GCC GGC ACA AGC CGC CGC AAG CTC CGT TCT      105
Ser Ala Asn Asn Lys Gly Ala Gly Thr Ser Arg Arg Lys Leu Arg Ser
   5                                10                                15
GAG AAG GCT GCA CTC CAG TTC TCC GTC AGT CGC GTC GAA TAC TCC CTC      153
Glu Lys Ala Ala Leu Gln Phe Ser Val Ser Arg Val Glu Tyr Ser Leu
  20                                25                                30                                35
AAG AAG GGG CGC TAT TGC AGG CGC TTA GGC GCT ACG GCC CCC GTC TAC      201
Lys Lys Gly Arg Tyr Cys Arg Arg Leu Gly Ala Thr Ala Pro Val Tyr
                                40                                45                                50
CTA GCC GCC GTC CTT GAA AAC CTC GTG GCC GAA GTG TTG GAC ATG GCG      249
Leu Ala Ala Val Leu Glu Asn Leu Val Ala Glu Val Leu Asp Met Ala
                                55                                60                                65
GCG AAC GTG ACA GAA GAA ACA TCC CCC ATT GTT ATC AAA CCG AGG CAT      297
Ala Asn Val Thr Glu Glu Thr Ser Pro Ile Val Ile Lys Pro Arg His
                                70                                75                                80
ATT ATG CTT GCC CCC AGG AAT GAT GTA GAA GTT GAA CAA GCT GTT TCA      345
Ile Met Leu Ala Pro Arg Asn Asp Val Glu Val Glu Gln Ala Val Ser
                                85                                90                                95
CGG TGT CAC CAT CTC GGC ATC AGG TGT CGT CCC TAAAACACGC AAAGAGCTGG      398
Arg Cys His His Leu Gly Ile Arg Cys Arg Pro
100                                105                                110
ACCGTCGCAA ACGCCGTTCC ACCTTTCAGC CGGATTAGTT CTTGATATTT CATTCTATCA      458
ATCTTGTTTA TGTGACTGTG ATTTTTCGTT TTGTGTTGAA CTAAGCCCCC TAATCTGGAT      518
TTCTCGTTTT ATGTTGAACT AAGTCTGTGC ACTCTTGAAG TAAAAAAAAA AAAAAAAAAA      578
AAAAAAAAAA

```

## (2) INFORMATION FOR SEQ ID NO:6:

- 25 -

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 110 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```

Met Ile Ser Ser Ala Asn Asn Lys Gly Ala Gly Thr Ser Arg Arg Lys
 1           5           10
Leu Arg Ser Glu Lys Ala Ala Leu Gln Phe Ser Val Ser Arg Val Glu
      20           25           30
Tyr Ser Leu Lys Lys Gly Arg Tyr Cys Arg Arg Leu Gly Ala Thr Ala
      35           40           45
Pro Val Tyr Leu Ala Ala Val Leu Glu Asn Leu Val Ala Glu Val Leu
      50           55           60
Asp Met Ala Ala Asn Val Thr Glu Glu Thr Ser Pro Ile Val Ile Lys
      65           70           75           80
Pro Arg His Ile Met Leu Ala Pro Arg Asn Asp Val Glu Val Glu Gln
      85           90           95
Ala Val Ser Arg Cys His His Leu Gly Ile Arg Cys Arg Pro
      100           105           110

```

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 485 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: CDS  
 (B) LOCATION: 16..348

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```

GATCCCAAAT CATCA ATG ACG ATC CCC GAA AAG AAA TCC GTC GCT CCG ATG      51
      Met Thr Ile Pro Glu Lys Lys Ser Val Ala Pro Met
      1           5           10
GCC CGT ATG AAG CAT ACA GCC CGC ATG TCT ACC GGC GGT AAG GCT CCA      99
Ala Arg Met Lys His Thr Ala Arg Met Ser Thr Gly Gly Lys Ala Pro
      15           20           25
CGC AAG CAG CTC GCC TCT AAG GCT CTT CGC AAG GCG CCA CCA CCA CCG      147
Arg Lys Gln Leu Ala Ser Lys Ala Leu Arg Lys Ala Pro Pro Pro Pro
      30           35           40
ACC AAA GGA GTG AAG CAG CCC ACC ACT ACC ACC TCC GGA AAA TGG CGC      195
Thr Lys Gly Val Lys Gln Pro Thr Thr Thr Thr Ser Gly Lys Trp Arg
      45           50           55           60
TTC GCG AGA TTT CAC AGG AAA CTG CCA TTC CAA GGG CTG GTG AGG AAA      243
Phe Ala Arg Phe His Arg Lys Leu Pro Phe Gln Gly Leu Val Arg Lys
      65           70           75

```

- 26 -

ATC TGG CAG GAC TTG AAG ACA CAT CTG CGC TTC AAG AAC CAC TCG GTT	291
Ile Trp Gln Asp Leu Lys Thr His Leu Arg Phe Lys Asn His Ser Val	
80 85 90	
CCT CCA CTT GAG GAG GTA ACT GAG GTT TAT CCT TGC CAA ACT ATT GGA	339
Pro Pro Leu Glu Glu Val Thr Glu Val Tyr Pro Cys Gln Thr Ile Gly	
95 100 105	
GGA TGC TAT TAGGATATTG AATTTGGATA ATGGTTTAAT TATCTGTTCT	388
Gly Cys Tyr	
110	
ACCTTTATGA TCAAATTTCT GTGGCTCAGC GTTGTGTAAT TTGGGCAATC GAATTCTTAG	448
CTATATTGCC TCAAAAAAAA AAAAAAAAAA AAAAAA	485

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 111 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Thr Ile Pro Glu Lys Lys Ser Val Ala Pro Met Ala Arg Met Lys	
1 5 10 15	
His Thr Ala Arg Met Ser Thr Gly Gly Lys Ala Pro Arg Lys Gln Leu	
20 25 30	
Ala Ser Lys Ala Leu Arg Lys Ala Pro Pro Pro Pro Thr Lys Gly Val	
35 40 45	
Lys Gln Pro Thr Thr Thr Thr Ser Gly Lys Trp Arg Phe Ala Arg Phe	
50 55 60	
His Arg Lys Leu Pro Phe Gln Gly Leu Val Arg Lys Ile Trp Gln Asp	
65 70 75 80	
Leu Lys Thr His Leu Arg Phe Lys Asn His Ser Val Pro Pro Leu Glu	
85 90 95	
Glu Val Thr Glu Val Tyr Pro Cys Gln Thr Ile Gly Gly Cys Tyr	
100 105 110	

DATED this 25th day of July 1997

**The University of Melbourne**

By DAVIES COLLISON CAVE

Patent Attorneys for the Applicants

GCCATCCCAT CAACAGAAGG TTTAAGTGGG AATCCATTTC ATTAGAAAAG ATCGGACAAA	60
GGGTACTCTT AAGCATACAA C ATG AGG GCG GTG GCG GTT TTC TTT GCT TGC	111
Met Arg Ala Val Ala Val Phe Phe Ala Cys	
1 5 10	
GTT CTC TTC TGT ATG GTT CAC AAA GCC GCA CTT GCG GAT GAT AAA ACG	159
Val Leu Phe Cys Met Val His Lys Ala Ala Leu Ala Asp Asp Lys Thr	
15 20 25	
TGC AAC CCT ACA GAT TTT ATG GTT ACC CAA ACC ATA ACT GGA TTG ACA	207
Cys Asn Pro Thr Asp Phe Met Val Thr Gln Thr Ile Thr Gly Leu Thr	
30 35 40	
ATC GGC GGT AAA CAA GAG TTC GAG GTC AAT TTA ATA AAC AAT TTG TAT	255
Ile Gly Gly Lys Gln Glu Phe Glu Val Asn Leu Ile Asn Asn Leu Tyr	
45 50 55	
TGT GCA CAA TCT AAT GTC AAA GTT TCA TGT GAC GGG CTT CAT ACC ACC	303
Cys Ala Gln Ser Asn Val Lys Val Ser Cys Asp Gly Leu His Thr Thr	
60 65 70	
GAA CCA ATA GAT CCT CAC ATT ATC AGA CCA CTT AGT GAC GGA ACG AAC	351
Glu Pro Ile Asp Pro His Ile Ile Arg Pro Leu Ser Asp Gly Thr Asn	
75 80 85 90	
AAC TGC CTT GTC AAC AAT GGA GCG CCT ATT TCT CAT GCT ACT CTT GTA	399
Asn Cys Leu Val Asn Asn Gly Ala Pro Ile Ser His Ala Thr Leu Val	
95 100 105	
GCA TTC AAG TAT GCC TGG GAT GTT CCT CCA TCT TTC AGC ATC ATC AGC	447
Ala Phe Lys Tyr Ala Trp Asp Val Pro Pro Ser Phe Ser Ile Ile Ser	
110 115 120	
TCT GAT ATA AAT TGC TCC TAA GGAGAAA ATTCTAGTTG GCAGAGAATA	495
Ser Asp Ile Asn Cys Ser OCH	
125	
ATCATATAGT CTTTTTTACT GAGCTATTTA ATTTTTTCAA TTTTCACCAA TAAGATTATT	555
TTAATGGAAT GTTAATGTAT TAGAATTGAA AAATAAAAAA AAAAAAAAAA AAAAAAAAAA	615
AAAAAAAAAA	625

FIGURE 1

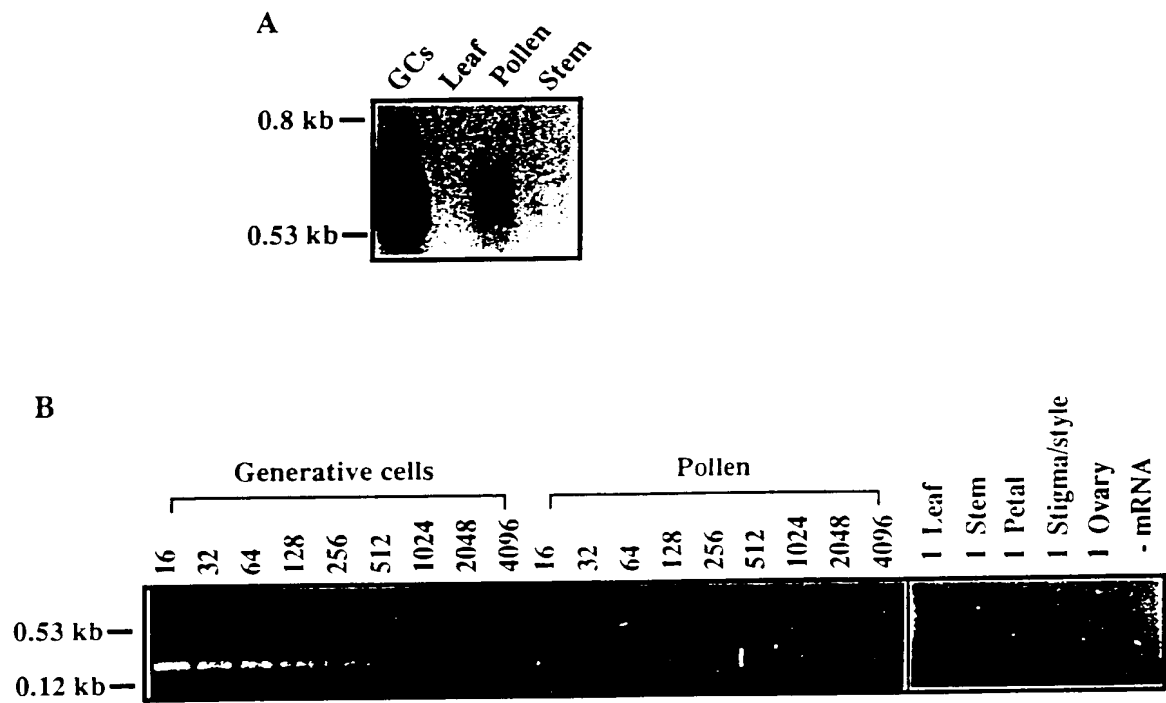
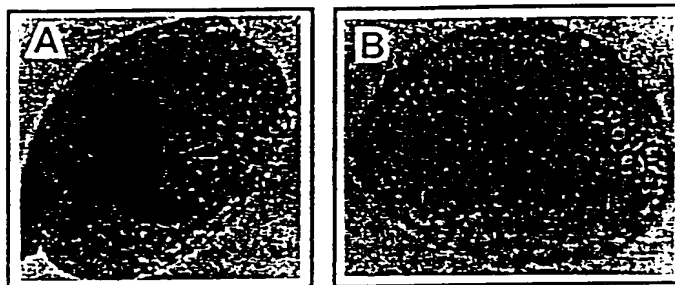


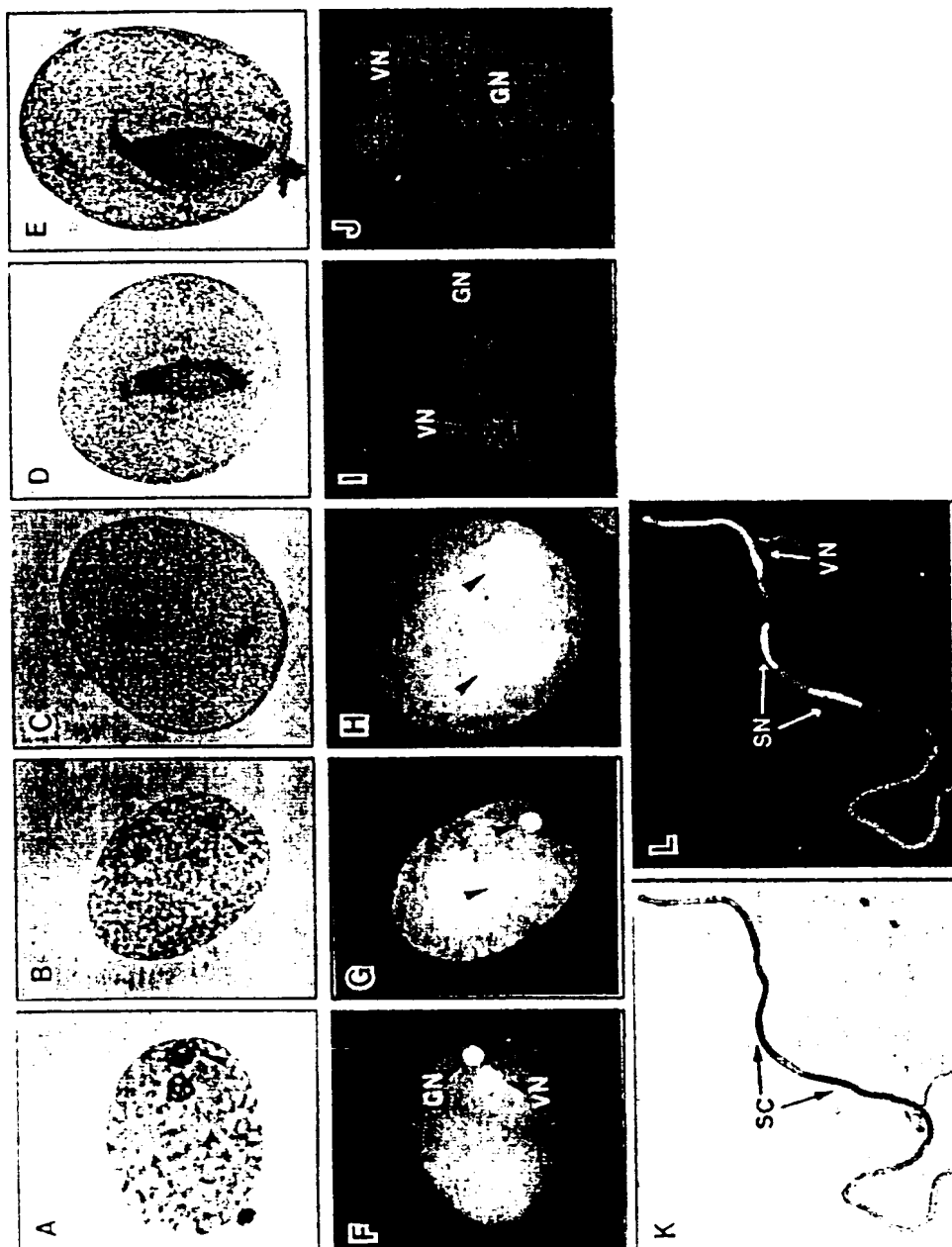
FIGURE 2

3/9



**FIGURE 3**

FIGURE 4





1 GAAAGTTGAAACATCTCCATCAAACCTCTAGAGTCAGATTTCACACAAG  
 49 ATGATTTCATCGGCAAATAACAAAGGCGCCGGCACAAGCCGCGCAAGCTCCGTTCTGAG 20  
 M I S S A N N K G A G T S R R K L R S E  
 109 AAGGCTGCACTCCAGTTCTCCGTCAGTCGCGTCGAATACTCCCTCAAGAAGGGGCGCTAT 40  
 K A A L Q F S V S R V E Y S L K K G R Y  
 169 TGCAGGCGCTTAGGCGCTACGGCCCCCGTCTACCTAGCCGCGTCCTTGAAAACCTCGTG 60  
 C R R L G A T A P V Y L A A V L E N L V  
 229 GCCGAAGTGTTGGACATGGCGGCGAACGTGACAGAAGAAACATCCCCCATTGTTATCAAA 80  
 A E V L D M A A N V T E E T S P I V I K  
 289 CCGAGGCATATTATGCTTGCCCCCAGGAATGATGTAGAAGTTGAACAAGCTGTTTCACGG 100  
 P R H I M L A P R N D V E V E Q A V S R  
 349 TGTCACCATCTCGGCATCAGGTGTCGTCCCTAAACACGCAAAGAGCTGGACCGTCGCAA 110  
 C H H L G I R C R P  
 409 ACGCCGTTCCACCTTTCAGCCGGATTAGTTCTTGATATTTTCATTCTATCAATCTTGGTTA  
 469 TGTGACTGTGATTTTTCGTTTTGTGTTGAACTAAGCCCCCTAATCTGGATTTCTCGTTTT  
 529 ATGTTGAACTAAGTCTGTGCACTCTTGAAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

FIGURE 5

1		GATCCCAAATCATCA	
16	ATGACGATCCCCGAAAAGAAATCCGTCGCTCCGATGGCCCGTATGAAGCATACAGCCCGC		
	M T I P E K K S V A P M A R M K H T A R	20	
76	ATGTCTACCGGCGGTAAGGCTCCACGCAAGCAGCTCGCCTCTAAGGCTCTTCGCAAGGCG		
	M S T G G K A P R K Q L A S K A L R K A	40	
136	CCACCACCACCGACCAAAGGASTGAAGCAGCCCACCACTACCACCTCCGGAAAATGGCGC		
	P P P P T K G V K Q P T T T T S G K W R	60	
196	TTCGCGAGATTTTACAGGAAACTGCCATTCCAAGGGCTGGTGAGGAAAATCTGGCAGGAC		
	F A R F H R K L P F Q G L V R K I W Q D	80	
256	TTGAAGACACATCTGCGCTTCAAGAACCACTCGGTTCCCTCCACTTGAGGAGGTAACCTGAG		
	L K T H L R F K N H S V P P L E E V T E	100	
316	GTTTATCCTTGCCAAACTATTGGAGGATGCTATTAGGATATTGAATTGGATAATGGTTT		
	V Y P C Q T I G G C Y	111	
376	AATTATCTGTTCTACCTTTATGATCAAATTTCTGTGGCTCAGCGTTGTGTAATTTGGGCA		
436	ATCGAATTCTTAGCTATATTGCCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA		

FIGURE 6

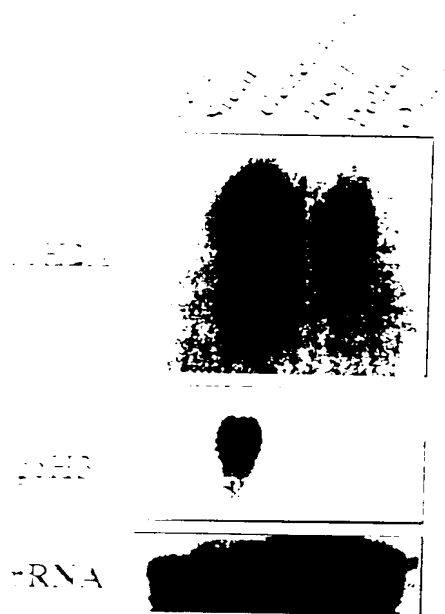


FIGURE 7

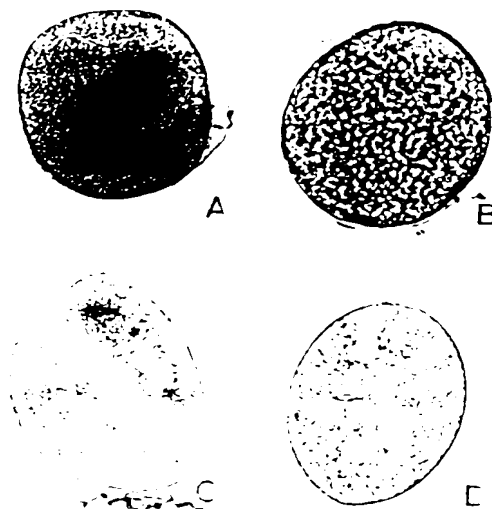


FIGURE 8

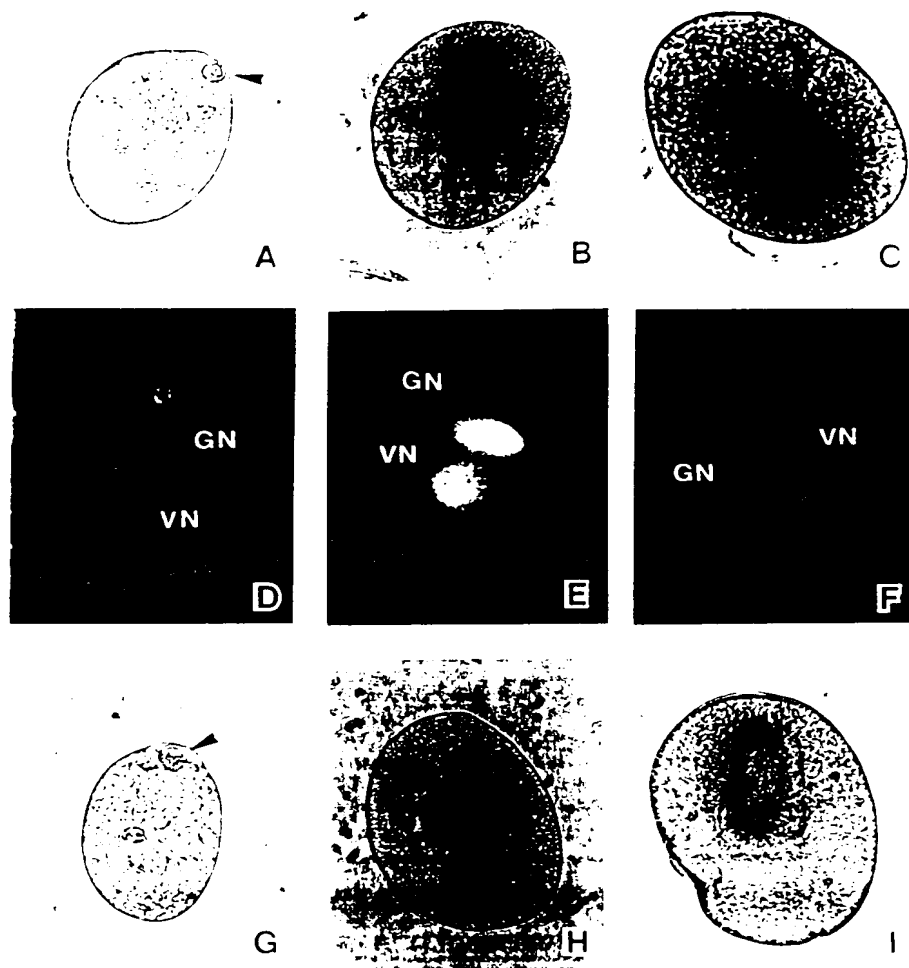


FIGURE 9

